Biomedical Signal Processing and Control 21 (2015) 146–156

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Ensemble artificial neural networks applied to predict the key risk factors of hip bone fracture for elders

Quan Liu [a,1], Xingran Cui [a,b,1], Yuan-Chao Chou [c], Maysam F. Abbod [d], Jinn Lin [e], Jiann-Shing Shieh [c,f,g,*]

[a] School of Information Engineering, Wuhan University of Technology, Wuhan, Hubei, China
[b] Department of Medicine, Beth Israel Deaconess Medical Center/Harvard Medical School, Boston, MA, USA
[c] Department of Mechanical Engineering, Yuan Ze University, Chung-Li, Taiwan
[d] College of Engineering, Design and Physical Sciences, Brunel University, London, Uxbridge UB8 3PH, United Kingdom
[e] Department of Orthopaedic Surgery, National Taiwan University Hospital, Taipei, Taiwan
[f] Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Chung-Li, Taiwan
[g] Center for Dynamical Biomarkers and Translational Medicine, National Central University, Chung-Li, Taiwan

ARTICLE INFO

Article history:
Received 23 September 2014
Received in revised form 10 May 2015
Accepted 4 June 2015
Available online 25 June 2015

Keywords:
Ensemble artificial neural networks
Back-propagation neural networks
Sensitivity analysis
Connection weights
Hip fracture

ABSTRACT

Hip bone fracture is one of the most important causes of morbidity and mortality in the elder adults. It is necessary to establish a prediction model to provide suggestions for elders. A total of 725 subjects were involved, including 228 patients with first low-trauma hip fracture and 497 ages-, sex-, and living area-matched controls (215 from the same hospital and 282 from community). All the subjects were interviewed with the same questionnaire, and the answers of the interviewees were recorded to the database. Three-layer back-propagation Artificial Neural Networks (ANN) models were applied for females and males separately in this study to predict the risk of hip bone fracture for elders. Furthermore, to improve the accuracies and the generalizations of the models, the ensemble ANNs method was applied. To understand variables contributions and find the important variables for predicting hip fracture, sensitivity analysis and connection weights approach were applied. In this study, three ANNs prediction models were tested with different architectures. With the fivefold cross-validation method evaluating the performances, one of the three models turned out to be the best prediction model and achieved a big success of prediction. The best area under the receiver operating characteristic (ROC) curve and the accuracy of the prediction model are 0.91 ± 0.028 (mean $\pm$ SD) and 0.85 ± 0.029 for females, while for males are 0.99 ± 0.015 and 0.93 ± 0.020. With the method of sensitivity analysis and connection weights, input variables were ranked according to contributions/importance, and the top 10 variables show great proportion of contribution to predict hip fracture. The top 10 important variables causing hip fracture for both females and males are similar to our previous results got from logistic regression model and other related researches. In conclusion, ANNs has successfully been used to establish prediction models for predicting the risk of hip bone fracture for both female and male elder adults respectively and identified the top 10 important variables from 74 input variables to predict hip bone fracture of elders. This study verified the performance of ANNs to be a highly efficient prediction model.

1. Introduction

Hip fracture is a kind of serious injury for elders. Previous studies have shown that elder adults with hip bone fracture have a relatively higher risk of death [1,2]. The post-fracture one-year mortality rates for the elders with hip fracture are 18–33% [3]. Even if the patients survive after the fracture, some of them still suffer functional loss in daily activities [4]. Moreover, the elders with hip fracture and their family need to shoulder much higher health care costs compared with their matched controls [5]. Therefore, hip fracture is not only a considerable health burden but also an increasing economic burden.

To reduce the incidence of this preventable injury and subsequent adverse outcomes, many studies have identified the risk

* Corresponding author at: Department of Mechanical Engineering, and Innovation Center for Big Data and Digital Convergence, Yuan Ze University, 135 Yuan-Tung Road, Chung-Li 32003, Taiwan. Tel.: +886 3 4638800 ext 2470.
E-mail address: jsshieh@saturn.yzu.edu.tw (J.-S. Shieh).
[1] These authors contributed equally to the work.

http://dx.doi.org/10.1016/j.bspc.2015.06.002
1746-8094/

factors for hip fracture [6,7,8]. Bone mineral density is considered to be the standard measure for the diagnosis of osteoporosis and the assessment of fracture risk [9]. Over the last decades, bone density testing has become widespread; yet the hope for a simple, straight predictor of fracture has faded. It is realized that the answer lies not in any single factor like bone density, but rather in assessment of multiple risk factors. Further, the operative risk factors may vary somewhat by populations/ethnicity studied. It is assumed that multi-variable risk assessments are indeed the tools of the future, and bone strength is determined by many factors, not just bone density; therefore many factors contribute to bone weakness. Recently, a matched case-control study carried out a conditional logistic regression to find out the important risk factors with the combined effects of different risk factors [10]. Artificial neural networks (ANNS) are another analysis method suitable for biomedical systems. According to the advantages of nonlinearity, fault tolerance, universality, and real-time operation, ANNs have been proposed as a quite suitable algorithm for modeling complex non-linear relationships in health care research [11,12,13]. Eller-Vainicher et al. [14] identified the promising role of ANN in predicting osteoporotic fracture among postmenopause osteoporosis women. For the comparison of the characteristics between ANNs and logistic regression applied to this epidemiological research field, a study has established prediction models for predicting living setting after hip fracture by ANNs and logistic regression, and shown that ANN is slightly better than logistic regression [15]. Lin et al. found ANN algorithm could reliably predict the mortality of hip fractured patients and outperforms the logistic regression method [16]. Although in many studies ANNs have been shown to exhibit superior predictive power compared to traditional approaches [17,18], they have also been labeled a "black box" because they provide little explanatory insight into the relative influence of the independent variables in the prediction process. This lack of explanatory power is a major concern to researchers since the interpretation of statistical models is desirable for gaining knowledge of the causal relationships. Besides, the significant ranking of each input is very important for the neural network operation. To "illuminate" the "black box", Olden et al. [19] introduced nine methods for quantifying variable importance in artificial neural networks, of which, sensitivity analysis is a generally used method. The sensitivity analysis methodology is able to show the specific contribution of the input variables while ANN has the capability to handle non-linear, complex ecological data and to incorporate causality [20,21]. Hence, the present study had two primary goals. The first goal was to establish ANNs prediction models to predict the risk of hip fracture for female and male elder adults respectively, and examine them via the ROC curve analysis. With this ANNs models, the second goal of this paper was to use the methods of sensitivity analysis and connection weights to understand the contribution of each input variable and identify the top 10 important variables for predicting hip fracture. These top 10 important variables were also compared with the most influential variables got from conventional logistic regression method [10,22].

2. Materials and methods

2.1. Database

The data used in this study was collected in the previous case-control matched study for the analysis of risk factors of hip fracture for elder adults aged 60 and older [10]. The data was collected using questionnaire surveys gathered by trained interviewers. The database included a total of 725 subjects, of which, 228 subjects were the patients admitted to the National Taiwan University Hospital with first low-trauma hip fracture, 215 subjects were hospital controls (patients in the same hospital but without hip fracture) and 282 subjects were community controls (randomly selected dwellers) individually matched to the hip fracture patients by age, gender, and living area, and then two control groups were combined together as 497 controls. Since women may have some risk factors different from men, such as reproductive history, etc, female and male models were developed separately, so the data was separated by gender. Of the total 725 subjects, 163 hip fracture patients and 345 controls were women, and 65 hip fracture patients and 152 controls were men. Moreover, in Lan's study [10], they used intraclass correlation coefficient to examine the reliability of the sample data. As a result, the moderate to high agreement suggested that the data was reliable.

2.2. Architecture of ensemble ANNs

There are many types of ANNs with different structures. Typical back-propagation neural networks (BPNN) are commonly adopted for solving classification problems. BPNN include an input layer, a hidden layer (or several hidden layers), and an output layer. Each layer contains at least 1 node (neuron). Activation functions

Table 1
The risk factors (input variables) for hip fracture analyzed in this study.

All risk factors in the questionnaire			
Ethnicity	Use of anti-diabetic	BMI	Height
Education	Use of analgesic	Cigarette smoking	Weight
Occupation	Use of other cardiovascular	Self-assessed health-comparison with same-aged people	Self-assessed health-comparison with one year ago
Marital status	Use of hypnotics	Leisure-time physical activity	Alcohol consumption
Monthly income	History of fall-induced fracture	Hormone replacement therapy	Number of children ever born
Living arrangement	History of fall	Intake of milk	Abortions or stillbirths
Hypertension	Bone fracture location	Intake of coffee	Pregnancy
Diabetes	History of fall at home	Intake of tea	Type of vegetarian diet
Stroke	History of fall outdoors	Intake of calcium	Total BMD value
Heart disease	Building type	Intake of vitamin	Use of walking aids
Parkinson's disease	Floor number where lived	Intake of glucosamine	Pain at walking
Arthritis	Stair lighting	Use of anti-hypertensive	Urinary incontinence
Osteoporosis	Stair height	MMSE score	Fecal incontinence
Liver disease	Number of stairs in a flight	Peak expiratory flow rate	Vision
Cancer	Outdoor lighting	Average hand grip strength	Headache or migraine
Cataract	Green light duration	Coordination	Hearing
Chronic respiratory disease	ADL difficulty	Menarche age	Self-assessed health-current
Constipation	IADL difficulty	Menopause age	
Weakness	Mobility difficulty	Duration between menarche and menopause	